Fluctuation scaling and covariance matrix of constituents' flows on a bipartite graph

AUTHOR(S):

Sato, Aki-Hiro; Hayashi, Takaki

CITATION:

Sato, Aki-Hiro ...[et al]. Fluctuation scaling and covariance matrix of constituents' flows on a bipartite graph. The European Physical Journal B 2010, 76(4): 529-535

ISSUE DATE:

2010-08

URL:

http://hdl.handle.net/2433/130801

Fluctuation scaling and covariance matrix of constituents' flows on a bipartite graph: Empirical analysis with high-frequency financial data based on a Poisson mixture model

Aki-Hiro Sato[1], Takaki Hayashi[2]
[1] Department of Applied Mathematics and Physics,
Graduate School of Informatics,
Kyoto University, Kyoto 606-8501, Japan

[2] Graduate School of Business Administration,
Keio University, 4-1-1 Hiyoshi Kouhoku-ku,
Yokohama Kanagawa, 223-8526, JAPAN

Abstract

We investigate a power-law relationship (means of constituents' flows versus their standard deviations) and variance/covariance matrix on a directed bipartite network. We propose a Poisson mixture model and a method to infer states of the constituents' flows on such a bipartite network from empirical observation without *a priori* knowledge on the network structure. By using a proposed parameter estimation method with high frequency financial data we found that the scaling exponent and simultaneous cross-correlation matrix have a positive correspondence relationship. Consequently we conclude that the scaling exponent tends to be 1/2 in the case of desynchronous (specific dynamics is dominant), and to be 1 in the case of synchronous (common dynamics is dominant).
PACS: 64.60.aq Networks; 89.65.Gh Economics, econophysics, financial markets; 02.50.-r Probability theory, stochastic processes, and statistics

1 Introduction

Recent accumulation of massive data in a broad spectrum of fields in our modern society requires to develop mathematical methodology to handle such large-scale data and to extract meaningful information from them, with which we need to make various decisions under uncertainty. From this perspective it is important to find summary quantities to measure association among them. It is also important to pursue methods to separate "global" (or "common") trend and "local" (or "specific") fluctuations.

Recently several researchers pay remarkable attention to a power-law relationship between a mean of flows and their standard deviation on a network [1, 2]. The power law relationship between a mean of constituents' flows at each element and their standard deviation, (a fluctuation at each element) $\approx c\times$ (an average at each element)$^\alpha$ (both c and α are identical on all the elements) has been found in a wide range of sciences [1, 2, 3, 4]. This phenomenon was called "fluctuation scaling" or "Taylor's power law" in studies on natural populations [3, 4] and on networks where elements serve as transporters of constituents among them on the basis of various dynamic processes [1, 5, 6].

Menezes and Barabási, and Eisler, Bartos, and Kertész investigated the scaling law between temporal averages of the network's traffic and its temporal fluctuations (characterized by standard deviations) measured at different nodes for various real systems [1, 2]. They reported that various types of systems such as packet transfer on the Internet (computers and packets), mass transfer by chemical reactions in cells (cells and chemical agents), people or goods transfer on a traffic system (transporters and people or goods), show this scaling relation and that the scaling exponent α corresponds to characteristics of the system's dynamics. They further characterized global (external) fluctuation and specific (internal) fluctuations under the several assumptions based on random walks on a network.

The fluctuation scaling is empirically observed in the following manner. Let $X_{j,\Delta t}(k)$ ($j = 1, \ldots, M$ and $k = 0, 1, 2, \ldots, Q-1$) denote the number of constituents arrived at the j-th node for the k-th sampling period between $k\Delta t$ and $(k+1)\Delta t$, where $\Delta t (> 0)$ denotes the size of a given time window to count the number of constituents. From the mean (time average) of $X_{j,\Delta t}(k)$

$$\langle X_{j,\Delta t}\rangle = \frac{1}{Q}\sum_{k=0}^{Q-1} X_{j,\Delta t}(k), \tag{1}$$

and the standard deviation

$$\sigma_{j,\Delta t} = \sqrt{\frac{1}{Q}\sum_{k=0}^{Q-1}\Big(X_{j,\Delta t}(k) - \langle X_{j,\Delta t}\rangle\Big)^2}, \tag{2}$$

we can confirm a (cross-sectional) power-law relationship

$$\sigma_{j,\Delta t} = c\langle X_{j,\Delta t}\rangle^\alpha \quad (j = 1, \ldots, M), \tag{3}$$

where c is a positive constant, and $\alpha \quad (1/2 \le \alpha \le 1)$ denotes a scaling exponent. If α takes $1/2$, then the arrival of the constituents on each node is independently random, like independent Poisson random variables. On the other hand if α takes 1, then their arrival on each node is synchronous [1]. In other words, if $\alpha = 1/2$, the specific fluctuations which affect the individual elements that constitute the system are dominant. On the other hand, if $\alpha = 1$ the common fluctuations dominate the dynamics of each element.

Specifically in social and economic activities, if we regard the situation where people exchange goods, money or information with each other as a network, then it seems to be fruitful to study such social or economic systems from a network point of view. There are several empirical studies on money flows with measuring scaling exponents in financial markets [7, 8, 9, 10]. Several networks relating to human activity can be described as bipartite graphs with two kinds of nodes. For example, financial markets (financial commodities and participants), blog systems (blogs and bloggers), and economic systems (firms and goods/consumers) can be described as a directed bipartite graph [11, 12, 13, 14].

In the meantime, with comprehensive, high resolution data which records the behaviors of market participants we can infer their mutual relationships, in particular in the form of cross-covariance /correlation matrices with a high resolution. Several methods to estimate covariance matrices for such high resolution data with a high degree of accuracy have been considered [15, 16]. In addition, a good model designed for describing multi-dimensional market movements from a broader perspective should not only be able to reproduce the statistical properties (means, cross-correlations, and so forth) observed in the historical data streams more accurately but also facilitate our inferring the states of the market better.

In the present article, we particularly focus on a directed bipartite network on financial markets and investigate constituents' flows on such a network under the assumption that we can observe constituents arrived at each node. The aims of this article are to establish a methodology to infer the total states of constituents' flows on the network from empirical observation without knowledge on the network structure and to examine a relationship between fluctuation scaling and covariance matrix.

We propose a Poisson model with stochastic intensity to describe constituents' generation on the bipartite network, and attempt to separately infer the parameters which come from a common latent information arrival and specific latent information arrivals. Through empirical analysis we will check the validity of the proposed model and the parameter estimation procedure by comparing empirical results and theoretical ones.

2 Model and parameter estimation procedure

Consider a bipartite graph with M groups ($j = 1, \cdots, M$) and N_j participants belonging to the j-th group as shown in Fig. 1. Ignoring the birth-death process of participants the population of each groups N_j is assumed to be constant, so that the total population of this system N is fixed ($N = \sum_{j=1}^{M} N_j$).

We further suppose that each participant in the j-th group can create a kind of message with probability $q + q_j$ in $[(t-1)\delta, t\delta]$. q_j and q represent *specific* and *global* probabilities of which are random variables sampled from $G_j(q_j)$ and $G(q)$. Let n_j be the number of message creations at the j-th group in $[(k-1)\Delta t, k\Delta t]$. Then, under the assumption that N is sufficiently large and

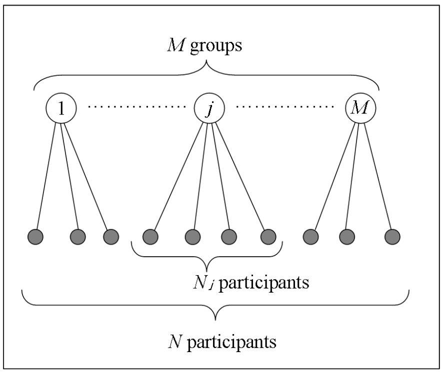


Figure 1: A schematic illustration of a bipartite representation for M groups consisting of $N_i \quad (i = 1, \ldots, M)$ participants.

$q + q_i$ is sufficiently small, n_j is approximated as a Poisson random variable with stochastic intensity $SN_j(q + q_j)$ [1] and its distribution and joint distribution are described as

$$\begin{aligned} F_j(n_j) &= \int_0^1 dq \int_0^1 dq_j G(q) G_j(q_j) \\ &\times \frac{SN_j(q + q_j)}{n_j!} e^{-SN_j(q+q_j)}, \end{aligned} \tag{4}$$

$$\begin{aligned} F_{lm}(n_l, n_m) &= \int_0^1 dq \int_0^1 dq_l \int_0^1 dq_m G(q) G_l(q_l) G_m(q_m) \\ &\times \frac{S^2 N_l(q + q_l) N_m(q + q_m)}{n_l! n_m!} \\ &\times e^{-SN_l(q+q_l) - SN_m(q+q_m)}, \end{aligned} \tag{5}$$

where S is the maximum integer less than $\Delta t/\delta t$. From Eqs. (4) and (5) one obtains

$$\langle n_j \rangle = SN_j(\langle q \rangle + \langle q_j \rangle), \tag{6}$$

$$\mathrm{Cov}(n_l, n_m) = \begin{cases} S^2 N_l N_m \sigma_q^2 & (l \neq m) \\ SN_l(\langle q \rangle + \langle q_l \rangle) + \\ \quad S^2 N_l^2 (\sigma_q^2 + \sigma_{q_l}^2) & (l = m) \end{cases}, \tag{7}$$

where $\langle q \rangle = \int_0^1 qG(q)dq$, $\langle q_l \rangle = \int_0^1 q_l G_l(q_l) dq_l$, $\sigma_q^2 = \int_0^1 (q - \langle q \rangle)^2 G(q) dq$, and $\sigma_{q_l}^2 = \int_0^1 (q_l - \langle q_l \rangle)^2 G_l(q_l) dq_l$. From Eqs. (6) and (7) we can reproduce the scaling exponent α, $\langle n_j \rangle$, and $\mathrm{Cov}(n_l, n_m)$ by determining $2M + 2$ parameters adequately.

[1] It is possible to derive the Poisson approximation from an agent-based model (See. App. A).

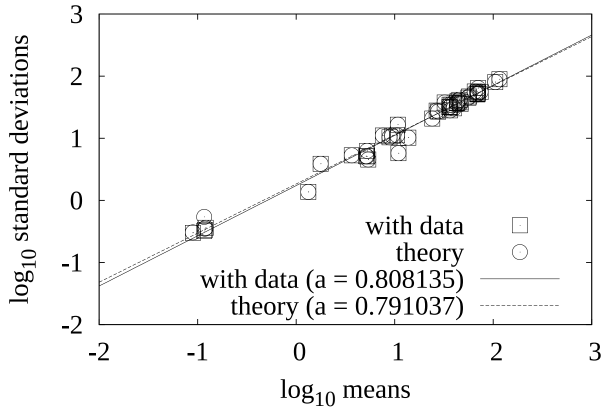


Figure 2: Log-log plots of the mean of quotation activities and their standard deviation for each currency pair obtained from empirical data (unfilled squares) at $\Delta t = 1$ [min] on 4th June 2007 for a period from 16:00 to 23:59 [UTC+2], and of the estimated mean and their standard deviation from the proposed model with the estimated parameters (unfilled circles). The x-axis represents the mean value, and the y-axis the standard deviation. The rms residuals of the empirical power-law relation ($\alpha(s) = 0.808135$) is 0.0872, and of the model-based estimation ($\alpha(s) = 0.791037$) 0.09438.

Let $X_{j,\Delta t}(k)$ be observed multiple time series for the numbers of message creations in $[(k-1)\Delta t, k\Delta t]$. Although the total number of participants N and the number of participants at each group can not be estimated from the observed time series, the relative existence ratio $K_j = (N_j/N)$ of participants belonging to the j-th group (degree centrality of the j-th group on a bipartite network as shown in Fig. 1) may be estimated as [17, 18]

$$K_j = \frac{\sum_{k=0}^{Q-1} X_{j,\Delta t}(k)}{\sum_{j=1}^{M} \sum_{k=0}^{Q-1} X_{j,\Delta t}(k)}. \tag{8}$$

Furthermore we can estimate windowed averages of the number of transactions/quotations and their covariance matrix in $[s\Delta t, (s+R-1)\Delta t]$ $(s = 0, \ldots, Q-R)$ with an overlapped time window having the length $R\Delta t$ $(0 < R < Q)$ as

$$\langle X_{j,\Delta t}\rangle^{\mathrm{emp}}(s) = \frac{1}{R}\sum_{k=s}^{s+R-1} X_{j,\Delta t}(k), \tag{9}$$

$$\begin{aligned} \mathrm{Cov}^{\mathrm{emp}}_{lm,\Delta t}(s) &= \frac{1}{R}\sum_{k=s}^{s+R-1} \Big(X_{l,\Delta t}(k) - \langle X_{l,\Delta t}\rangle^{\mathrm{emp}}(s)\Big) \\ &\times \Big(X_{m,\Delta t}(k) - \langle X_{m,\Delta t}\rangle^{\mathrm{emp}}(s)\Big). \end{aligned} \tag{10}$$

In order to infer parameters $\langle q\rangle + \langle q_j\rangle$, σ_q^2, and $\sigma_{q_j}^2$ we set an evaluation function

$$\begin{aligned}
\epsilon^2 &= \sum_{j=1}^{M}\Big(\langle X_{j,\Delta t}\rangle^{\mathrm{emp}}(s) - K_j SN(\langle q\rangle + \langle q_j\rangle)\Big)^2 \\
&+ \sum_{l=1}^{M}\Big(\mathrm{Cov}^{\mathrm{emp}}_{ll,\Delta t}(s) - \{K_l SN(\langle q\rangle + \langle q_l\rangle) \\
&+ K_l^2 S^2 N^2(\sigma_q^2 + \sigma_{q_l}^2)\}\Big)^2 \\
&+ \sum_{l=1}^{M}\sum_{m=l+1}^{M}\Big(\mathrm{Cov}^{\mathrm{emp}}_{lm,\Delta t}(s) - K_l K_m S^2 N^2 \sigma_q^2\Big)^2 \qquad (11)
\end{aligned}$$

and minimize ϵ^2 under the condition where $\langle q\rangle + \langle q_j\rangle \geq 0$, $\sigma_q^2 \geq 0$, and $\sigma_{q_j}^2 \geq 0$. Partially differentiating Eq. (11) by $\langle q\rangle + \langle q_j\rangle \quad (j = 1, \ldots, M)$, σ_q^2, and $\sigma_{q_j}^2 \quad (j = 1, \ldots, M)$ and setting them into zero, we have

$$\Big(\langle q\rangle + \langle q_j\rangle\Big)(s) = \frac{\langle X_{j,\Delta t}\rangle^{\mathrm{emp}}(s)}{K_j SN}, \qquad (12)$$

$$\sigma_q^2(s) = \frac{\sum_{l=1}^{M}\sum_{m=l+1}^{M}\mathrm{Cov}^{\mathrm{emp}}_{lm,\Delta t}(s) K_l K_m}{S^2 N^2 \sum_{l=1}^{M}\sum_{m=l+1}^{M}(K_l K_m)^2}, \qquad (13)$$

$$\sigma_{q_j}^2(s) = \frac{\mathrm{Cov}^{\mathrm{emp}}_{jj,\Delta t}(s) - \langle X_{j,\Delta t}\rangle^{\mathrm{emp}}(s)}{K_j^2 S^2 N^2} - \sigma_q^2(s). \qquad (14)$$

Indeed, Eqs. (12) and (13) turn out natural, as we obtain $\langle X_{j,\Delta t}\rangle^{\mathrm{emp}}(s) = \langle n_j\rangle$ and $\mathrm{Cov}^{\mathrm{emp}}_{ll,\Delta t}(s) = \mathrm{Cov}(n_l, n_l)$ by using Eqs. (6), (7), (12) and (13).

3 Analysis

In order to verify both the proposed model and parameter estimation procedure we use the numbers of quotations extracted from the high-resolution data in the foreign exchange market [19]. By using such multivariate time series we calculate a *cross-sectional* relationship of the foreign exchange market, so that we compute the power-law exponents of fluctuation scaling and their covariance / correlation matrices by means of both empirical procedure by Eqs. (1) and (2) and model-based estimation procedure by Eqs. (6), (7), (8), (12), (13), and (14).

To do so we regard quotations which bank traders and brokers offer and take as messages (constituents moving on the bipartite network) in the proposed model, the traders and brokers as participants, and market places where they

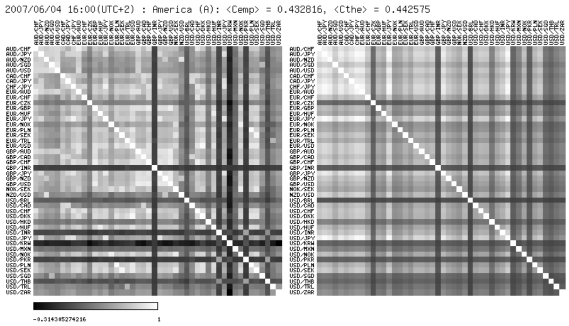


Figure 3: A cross-correlation matrix $C^{emp}_{ij,\Delta t}(s)$ obtained from empirical data at $\Delta t = 1$ [min] (left) and one obtained through the proposed model C_{ij} (right) for periods from 16:00 to 23:59 [UTC+2] on 4th June 2007.

exchange pairwise currencies as groups. Then we can assume that a quotation for every market participant to create in each market place corresponds to a linkage between the market place and the market participant. Namely, this bipartite network represents a *system* constructed by participants' perception and action.

Counting the number of quotations per $\Delta t = 1$ [min] we obtain multiple time series of quotations' flows in the foreign exchange market. First of all, we investigate a cross-sectional relationship of community with means of quotation activities and their standard deviations. Fig. 2 shows double-logarithmic scatter plots of mean and standard deviation obtained from quotation activities for each currency pair on an arbitrarily chosen period, from 16:00 to 23:59 on 4th June 2007 (s=960). Each unfilled square represents mean of quotation activities and their standard deviations for the period. From Fig. 2 we can find that quotation activities of the foreign exchange market follow the fluctuation scaling [20]. Regarding Eq. (3) by means of the least square method for a fitting function $(j = 1, \ldots, M)$

$$\log \mathrm{Cov}^{\mathrm{emp}}_{jj,\Delta t}(s) = 2\alpha(s) \log\langle X_{j,\Delta t}\rangle^{\mathrm{emp}}(s) + 2 \log c(s) \tag{15}$$

we obtained $(\log c(s), \alpha(s)) = (0.237899, 0.808135)$ at $s = 960$ ($R^2 = 0.9833$). Furthermore applying the least square method to other periods on 4th June 2007 we obtained, for instance, $(0.29757, 0.764023)$ at $s = 0$ (0:00-7:59) and $(0.208729, 0.772521)$ at $s = 480$ (8:00-15:59), respectively. Fig. 4 (a) shows the trajectory of $(\log c(s), \alpha(s)) \quad (s = 0, \ldots, 4999)$ from 00:00 on 4th June 2007 to 19:19 on 8th June 2007. From Fig. 4 (a) we can find that $\alpha(s)$ and $\log c(s)$ vary in sampling period s.

Next we calculated $\langle X_{j,\Delta t}\rangle^{\mathrm{emp}}$, $\mathrm{Cov}^{\mathrm{emp}}_{lm,\Delta t}$, and K_j with the multiple time series by setting $R\Delta t = 480$ [min] and $Q\Delta t = 7,200$ [min]. Substituting them into Eqs. (12), (13), and (14) we estimated values of $2M + 1$ parameters. We set a free parameter $SN = 500,000$ in order to prevent the smallest group population $\min_j\{M_j\}$ from vanishing. Note that we confirmed robustness of results for SN. Unfilled circles shown in Fig. 2 exhibit that it is found that the scaling relationship may be reproduced by Eqs. (6) and (7) with estimated parameters shown in Tab. 1. Furthermore Fig. 3 shows gray scale display of simultaneous cross-correlation matrices defined as

$$C_{ij} \equiv \frac{\mathrm{Cov}(n_i, n_j)}{\sqrt{\mathrm{Cov}(n_i, n_i)\mathrm{Cov}(n_j, n_j)}}, \tag{16}$$

$$C^{\mathrm{emp}}_{ij,\Delta t}(s) \equiv \frac{\mathrm{Cov}^{\mathrm{emp}}_{ij,\Delta t}(s)}{\sqrt{\mathrm{Cov}^{\mathrm{emp}}_{ii,\Delta t}(s)\mathrm{Cov}^{\mathrm{emp}}_{jj,\Delta t}(s)}}. \tag{17}$$

As shown in Fig. 3 the simultaneous cross-correlation matrix estimated by using Eq. (7) with estimated parameters captures a tendency of that obtained by using Eq. (10). Specifically cross-correlation matrices obtained from Eq. (7) are consistent with those from empirical data for currencies with small K_j.

Although the reproduction of fluctuation scaling by using the propose model and estimation procedure is relatively better than that by the model of random walkers on a network in the presence of external driving [2], the covariance matrices calculated from theoretical model are sometimes in disagreement with those estimated from empirical data. This gap may be caused by the simplistic assumptions regarding two types of information arrivals. Incorporating a hierarchical structure of information arrivals may be one potential direction to alleviate such discrepancies.

Moreover we examined a *temporal* relationship between the scaling exponent and global mean of simultaneous cross-correlations

$$\langle C^{\mathrm{emp}}_{\Delta t}\rangle(s) \equiv \frac{1}{2M(M-1)}\sum_{i=1}^{M}\sum_{j=i+1}^{M} C^{\mathrm{emp}}_{ij,\Delta t}(s). \tag{18}$$

We calculated the scaling exponent $\alpha(s)$ and the global mean of simultaneous cross-correlation matrix $\langle C^{\mathrm{emp}}_{\Delta t}\rangle(s)$ for each s from Eqs. (15) and (18). Fig. 4 (b) shows the trajectory of $(\alpha(s), \langle C^{\mathrm{emp}}_{\Delta t}\rangle(s))$ estimated from the data from 4th till 8th of June 2007. From it we found roughly a linear relationship with positive slope between α and the overall mean of the simultaneous cross-correlation matrix. Since α takes a non-trivial value at $\Delta t = 1$ [min], the quotation activities of the foreign exchange market may be mutually correlated due to internal and external factors or strong temporal correlations like inhomogeneity of Hurst exponents [2]. In Ref. [20] since one of the author (AHS) shows that the scaling exponent α increases as Δt increases, the quotation activities may become more temporally correlated or mutually synchronous than small Δt. This implies that for a large time scale, the market participants synchronously behave and may be driven better by a common factor than specific factors. This is related to daily pattern of market participants.

Consequently, as an explanation in accordance with the proposed model the nontrivial values of α may imply that participants are affected by both information from different origins relating to specific (independent) factors (microscopic dynamics) and from sources relating to common (herding) factors (macroscopic dynamics). $2(M+1)$ parameters $\langle q \rangle$, $\langle q_j \rangle$, σ_q, and σ_{q_j} in the proposed model characterize such factors as both herding effects and independent randomness. As another explanation they may imply that the behavior of market participants has a strong temporal correlation.

We argue that the scaling exponent α is determined by participants' responses to specific information arrivals and common information arrivals. The ratio of $\sigma_{q_j}^2$ to σ_q^2 $\eta_j \equiv \sigma_{q_j}^2/\sigma_q^2$ can be an indicator to measure dominant factors which affect participants belonging to each group. If $\eta_j \gg (\ll) 1$ then the specific (common) information arrivals dominate participants' actions. Tab. 1 shows the estimated parameters $\langle q \rangle + \langle q_j \rangle$ and $\eta_j \equiv \sigma_{q_j}/\sigma_q$ by means of the proposed method and centrality K_j. Fig. 5 shows a relation between the centrality K_j and the ratio of specific probability fluctuation to a common probability fluctuation η_j. From Fig. 5 it is found that groups having large centrality K_j have a tendency to be affected by common information arrivals. Therefore the proposed model and parameter estimation procedure enable us to characterize global tendency and specific fluctuations from participants' activities of message creation.

4 Discussion

The proposed model can be found to almost completely reproduce the power-law relationship between the mean and their standard deviation with non-trivial exponent values $(1/2 < \alpha < 1)$.

Furthermore the proposed model can infer activities of which agents participate in currency exchanges at each currency pair of the foreign exchange market. Throughout the empirical investigation it is confirmed that cross-correlation among activities at each currency pairs show strong association with the scaling index of fluctuation scaling. In fact it substantially burdensome to calculate the total mean of cross-correlation matrix, while it is relatively facile to estimate the index of fluctuation scaling. The results of our model investigation imply that we can use the scaling exponent of fluctuation scaling as a summary index of the activities of the whole foreign exchange market.

In the literature of financial markets Geman *et al.* [21] proved that the Poisson model with reflected normal jumps' intensity can be constructed by a Poissonian time-change of a univariate Brownian motion. An attempt to construct multivariate subordination has been made by some authors [22]. Our proposed model was empirically found to be adequate for studying financial markets including the foreign exchange market. Based on the estimated values of the parameters for the proposed model and their evolution through time, financial assets can be discriminated and their behavioral tendencies can be captured. They could be utilized for prediction and trading purposes in the foreign exchange market.

Moreover in the context of finance the normal mixture hypothesis (or more generally "Mixture of distribution hypothesis") has been proposed as an alternative explanation for the description of return distribution of financial assets by several studies [23, 24, 25]. They have considered trading volume or the number of transactions (or quotations) as a proxy of the latent number of information arrivals. The model proposed in the present article may be extended to include return dynamics so as to jointly model return-volume relationship [26, 27, 28, 29].

5 Conclusion

We considered a bipartite model for message generation by N agents in M groups, and proposed the Poisson model with stochastic intensity for message creations with its parameter estimation procedure.

We approximately derived the model with stochastic intensity from the threshold agent-based model where the probabilities for agents to decide their actions randomly vary in time. It was shown that the stochastic intensity can be modelled as two parts originated from common information arrivals and specific information arrivals. It may be concluded that the estimated parameters characterize behavioral tendencies of participants driven by common information arrivals and specific information arrivals.

We investigated quotation activities in the foreign exchange market and estimated model parameters from the high-resolution data containing 45 currency pairs by means of the proposed method. Furthermore we found that there is an obvious association between the exponents of fluctuation scaling for the numbers of quotations per minute and their cross-correlation matrix. As the values of α are nontrivial ($1/2 < \alpha < 1$), it may quantify magnitude of influences which come from both common and specific factors from a holistic point of view. As another possible hypothesis Hurst exponent heterogeneity should be considered.

The scaling analysis between means and standard deviations is among useful methods to understand the whole states of human activities in a community. Further investigation with the proposed approach and its potential extensions applying to large-scale datasets would bring fruitful insights on the states of such a community in a broad spectrum of applications.

Various types of physical aspects can contribute to make deeper understanding of herding behavior of heterogeneous agents in social and economic systems.

Acknowledgement

One of the authors (A.-H. S) appreciates valuable comments and suggestions by Prof. T. Munakata, Assoc. Prof. A. Igarashi, and Prof. J. A. Hołyst. This work was partially supported by the Grant-in-Aid for Young Scientists (B) (21760059) of the Japanese Ministry of Education, Culture, Sports, Science and Technology (A.-H. Sato). Furthermore A.-H. Sato is grateful for the financial support from Ken Millennium, Corporation. T. Hayashi acknowledges partial financial support from the Grant-in-Aid for Scientific Research (C) (19530186), Japan Society for the Promotion of Science.

A Derivation of the Poisson model with random intensities from an agent-based model

Here we derive the compound Poisson processes proposed in Sec. 2 from a stochastic agent-based model [13].

Consider a bipartite graph with two kinds of objects, M groups ($j = 1, \cdots, M$) and N participants ($i = 1, \cdots, N$) as shown in Fig. 1. We suppose that the participants in a group can create a kind of message based on the double-threshold agent model [13], and assign N_j participants ($\sum_{j=1}^{M} N_j = N$) to the j-th group. In the context of financial markets a message corresponds to a quotation and a group to a board of a financial commodity, and of blog systems a message to a document and a group to a content.

Every participant can take two types of actions, coded as 1 (creating a message) and 0 (not responding). Let $y_i^j(t)$ denote an action of the i-th participant who attends the j-th group during a period $[t\delta, (t+1)\delta] \quad (t = 0, 1, \ldots)$, which takes either 0 or 1. δ represents the shortest time interval within which participants can decide one action to take, and Δt represents the time length for regular observation intervals. The probabilities for the i-th agent in the j-th group to be 0 or 1 are given by

$$\mathrm{Prob}\Big[y_i^j(t) = 1\Big] = p_{i,j}(t), \tag{19}$$

$$\mathrm{Prob}\Big[y_i^j(t) = 0\Big] = 1 - p_{i,j}(t), \tag{20}$$

where $p_{i,j}(t)$ are random variables taking a value from 0 to 1.

According to [13, 14, 30], the participant determines his/her action based on an input (perceived information). Therefore we could assume that $p_{i,j}(t)$ is described as $p_{i,j}(t) = Q_{i,j}\Big(x_i^j(t)\Big)$, where $Q_{i,j}(x)$ is an S-curve function ($0 \leq Q_{i,j}(x) \leq 1$), and $x_i^j(t)$ denotes the information which the i-th participant in the j-th group perceives during the period t. If we assume that all the participants in the j-th group perceive the same information $x^j(t)$ and have an identical function form of $p_{i,j}(\cdot)$, then we can replace $p_{i,j}(t)$ as $p_j(t) \quad (0 \leq p_j(t) \leq 1)$. Then the number of messages created in the j-th group during the k-th sampling period is described as,

$$f_j^S(k) = \sum_{i=1}^{N_j} \sum_{t=kS}^{(k+1)S-1} y_i^j(t) \quad (k = 0, 1, \ldots, Q-1), \tag{21}$$

where S is the maximum integer less than $\Delta t/\delta$. Under a fixed p_j the probability distribution of Eq. (21) is given by a binomial distribution, so that

$$\begin{aligned} G_j(n_j|p_j) &\equiv \mathrm{Prob}[f_j^S(K) = n_j|p_j] \\ &= \begin{pmatrix} SN_j \\ n_j \end{pmatrix} {p_j}^{n_j}(1-p_j)^{SN_j - n_j} \quad (0 \leq n_j \leq SN_j), \end{aligned} \tag{22}$$

where $\binom{SN_j}{n_j}$ represent binomial coefficients.

For the sake of convenience we suppose the following situations: (a) The probability $p_j(t)$ is determined by specific information common to the participants belonging to the j-th group and information common to all the participants. It corresponds to a specific event in the system. Then we may express it as $p_j(t) = q_j(t) + q(t) - q_j(t)q(t)$, where $q_j(t)$ $(0 \leq q_j(t) \leq 1)$ represents the probability for participants belonging to the j-th group to response specific information, and $q(t)$ $(0 \leq q(t) \leq 1)$ to response common information. (b) The values of $q_j(t)$ and $q(t)$ are random variables sampled from the probability density function $F_j(q_j)$ and $F(q)$, respectively. $F_j(q_j)$ are probability density functions with the mean $\langle q_j \rangle$ and the variance ${\sigma_{q_j}}^2$, and $F(q)$ has the mean $\langle q \rangle$ and the variance ${\sigma_q}^2$. If we assume that q, q_l, and q_m are mutually independent and that SN is sufficiently large and q and q_j are sufficiently small, then we can use a Poisson approximation to the binomial distributions and approximate n_j as Poisson random variables with stochastic intensity $SN_j(q + q_j)$.

References

[1] M.A. de Menezes and A.-L. Barabási, Physical Review Letters, **92**, 028701 (2004)

[2] Z. Eisler, I. Bartos, and J. Kertész, Advances in Physics, **57**, 89 (2008)

[3] R.M. Anderson, D.M. Gordon, M.J. Crawley, and M.P. Hassell, Nature, **296**, 245 (1982)

[4] A.M. Kilpatrick, and A.R. Ives, Nature, **422**, 65 (2003)

[5] M.A. de Menezes and A.-L. Barabási, Physical Review Letters, **93**, 068701 (2004)

[6] J. Duch and A. Arenas, Physical Review Letters, **96**, 218702 (2006)

[7] Z. Eisler, J. Kertész, S.-H. Yook, and A.-L. Barabási, Europhysics Letters, **69**, 664 (2005)

[8] Z.-Q. Jiang, L. Guo, and W.-X. Zhou, European Physical Journal B, **57**, 347 (2007)

[9] A.-H. Sato, H. Sakai, M. Nishimura, and J.A. Hołyst, Progress of Theoretical Physics Supplement, **179**, 38 (2009)

[10] Z. Eisler and J. Kertész, Physical Review E, **73**, 046109 (2006)

[11] R. Lambiotte, M. Ausloos, and M. Thelwall, Journal of Informetrics, **1**, 277 (2007)

[12] A.M. Chmiel, J. Sienkiewicz, K. Suchecki, and J. A.Hołyst, Physica A, **383**, 134 (2007)

[13] A.-H. Sato, Physica A, **382**, 258 (2007)

[14] A.-H. Sato and J. A. Hołyst, European Physical Journal B, **62**, 373 (2008)

[15] M.M. Dacorogna, R. Gençay, U. Müller, R.B. Olsen, and O.V. Pictet, *An introduction to high-frequency finance*, (Academic Press, San Diego, 2000)

[16] T. Hayashi and N. Yoshida, Bernoulli, **11-2**, 359 (2005)

[17] S. Meloni, J. Gómez-Gardeñes, V. Latora, and Y. Moreno, Physical Review Letters, **100**, 208701 (2008)

[18] J.D. Noh and H. Rieger, Physical Review Letters, **92**, 118701 (2004)

[19] The data are provided by CQG Inc.

[20] A.-H. Sato, M. Nishimura, and J.A. Hołyst, Physica A, **389**, 2793 (2010)

[21] H. Geman, D.B. Madan, and M. Yor., Math. Finance, **11**, 79–96 (2001)

[22] E. Luciano, P. Semerao, Journal of Computational and Applied Mathematics, **233**, 1937–1953 (2010)

[23] P. Clark, Econometrica, **41**, 135 (1973)

[24] T. Tauchen, and M. Pitts, Econometric Society, **51**, 485 (1983)

[25] M. Richardson, and T. Smith, Journal of Financial and Quantitative Analysis, **29**, 101 (1994)

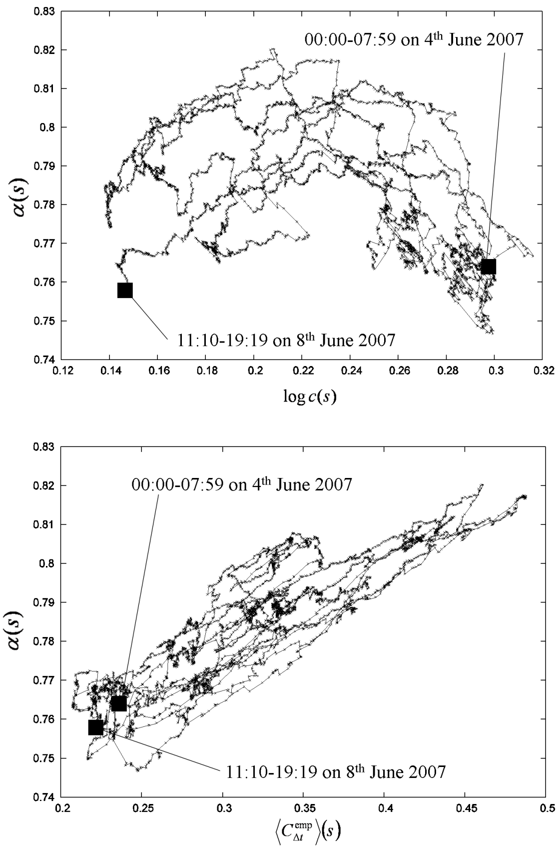
[26] C.G. Lamoureux, and W.D. Lastrapes, J. Finance, **45**, 221 (1990)

[27] T.G. Andersen, J. Finance, **51**, 169 (1996)

[28] R. Liesenfeld, Journal of Business and Economic Statistics, **16**, 101 (1998)

[29] T. Watanabe, Journal of Business and Economic Statistics, **18**, 199 (2000)

[30] A. Krawiecki, J.A. Hołyst, and D. Helbing, Physical Review Letters, **89**, 158701 (2002)



(a) (b)

Figure 4: The trajectory of scaling exponents and positive constants of fluctuation scaling $(\log c(s), \alpha(s))$ (a) and that of scaling exponents and global means of simultaneous cross-correlation matrix $(\langle C^{\mathrm{emp}}_{\Delta t}\rangle(s), \alpha(s))$ (b) for each s at $R\Delta t = 480$ [min]. These two trajectories are estimated from date for a period between 00:00-07:59 on 4th (s=0) and 11:19-19:19 on 8th June 2007 (s=4999).

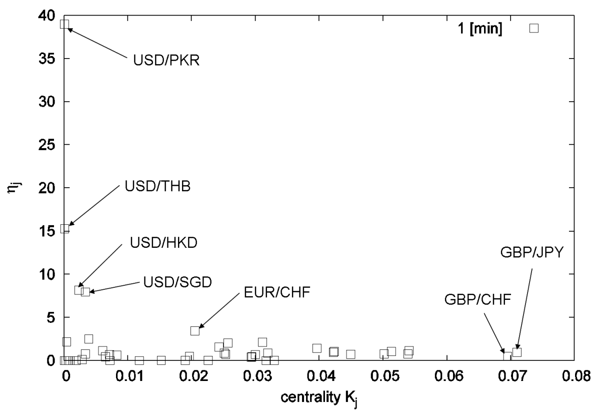


Figure 5: Relationship between centrality and the ratio of probability fluctuations (measured by their variance) between specific information arrivals and common information arrivals for each currency pair obtained from empirical data during a period from 4th to 8th June 2007 at $\Delta t = 1$ [min].

Table 1: Relationship between centrality and the ratio of probability fluctuations (measured by their variance) between specific information arrivals and common information arrivals for each currency pair obtained from empirical data during a period from 4th to 8th June 2007 at $\Delta t = 1$ [min]. The x-axis represents the value of centrality, and the y-axis the ratio of fluctuations. We set $SN = 500,000$ and estimated parameters during 00:00-7:59 [UTC+2] on 4th ($R\Delta t = 480$ [min]). We obtained the fluctuation from a common information arrivals $\sigma_q = 9.7038 \times 10^{-5}$ and the error $\epsilon^2 = 17,512,854.34$. We list them up in descending order of K_j

currency pairs	$\langle q \rangle + \langle q_j \rangle$ ($\times 10^{-5}$)	$\eta_j = \sigma_{q_j}^2 / \sigma_q^2$	K_j ($\times 10^{-4}$)
USD/PKR	25.698978	38.980106	0.061483
GBP/INR	30.440692	0.000000	0.062833
USD/THB	26.364087	15.260999	0.086743
USD/KRW	8.893520	0.000000	0.294545
USD/INR	14.720686	2.196643	0.409567
USD/BRL	5.914299	0.000000	0.653829
EUR/CZK	0.178804	0.000000	0.930183
USD/TRL	2.689206	0.000000	1.886342
USD/HKD	22.351278	8.160153	2.312351
EUR/TRL	6.055118	0.117646	2.794839
USD/HUF	12.767622	0.768017	3.357635
USD/SGD	15.718264	7.939419	3.364864
AUD/SGD	16.327196	2.475077	3.870944
EUR/PLN	12.347652	1.152112	6.014258
EUR/HUF	10.078562	0.422157	6.522562
USD/PLN	11.628551	0.691049	7.144220
USD/ZAR	1.249443	0.000000	7.188226
CAD/CHF	15.624402	0.618260	8.244473
USD/MXN	0.653940	0.000000	11.826591
GBP/NZD	10.922223	0.023758	15.231490
EUR/NOK	9.544064	0.030554	18.990590
USD/DKK	10.607196	0.513997	19.646961
EUR/CHF	21.541244	3.414166	20.507003
EUR/SEK	9.078926	0.005723	22.596936
EUR/GBP	15.843120	1.578475	24.289615
AUD/CHF	17.460789	0.856311	25.108828
GBP/CAD	11.553375	0.746318	25.340222
USD/CAD	10.487927	2.010160	25.686399
NZD/USD	9.730851	0.389707	29.338500
AUD/NZD	12.822906	0.430525	29.433425
GBP/AUD	14.202507	0.699970	29.949911
AUD/USD	16.304866	2.122383	31.134921
USD/NOK	8.364065	0.000000	31.706773
EUR/USD	12.966842	0.840223	31.902898
USD/SEK	9.095179	0.026573	32.957079
USD/JPY	15.842229	1.445519	39.660832
GBP/USD	10.295732	0.976898	42.227295
USD/CHF	14.770476	1.060830	42.355662
CAD/JPY	13.310597	0.754875	44.951914
EUR/JPY	17.379840	0.766174	50.157285
EUR/AUD	15.564179	1.044889	51.330619
CHF/JPY	17.891211	0.785571	53.878415
AUD/JPY	17.962393	1.151438	54.104725
GBP/CHF	14.337831	0.478344	69.516453
GBP/JPY	15.772803	0.950204	70.968762